Predicting the Price of Bitcoin, Dogecoin and Ethereum by Machine Learning

Siyi Liao *

Department of Finance, Southwestern University of Finance and Economics, Chengdu, China

*Corresponding author: 42004284@smail.swufe.edu.cn

Abstract. Contemporarily, blockchains and cryptocurrencies have gained their popularity among investors and hedge funder, where both have bright prospects. On this basis, cryptocurrencies have been used in trading more and more with the development of website and computer. In this case, their prices fluctuations do have great significance to the public. This paper chooses three machine learning model (i.e., XGBoost, LightGBM and Linear Model) to predict the price of three cryptocurrencies (i.e., Bitcoin, Dogecoin and Ethereum). To be specific, this study uses the data from 2020-01-01 to 2022-12-07, including close price, open price, high price, low price, and the volume of trading coins. According to the analysis, Linear Model can predict the price best, with well-fitted trend prediction and accurate price prediction. In addition, other models can also have good predictions but they are not better than Linear model. These results can help others to predict the price of cryptocurrencies and have a deep understanding of cryptocurrency and machine learning.

Keywords: Cryptocurrency; machine learning; prediction of price.

1. Introduction

Blockchains are one projects which are integrated into innovative technologies, based on many theoretical and practical successes. In 1976, Diffie and Hellman published the significant paper New directions in cryptography, covering the further steps of all aspects of branches of encryption, including the asymmetric encryption, elliptic curve and so on, making great effect on the development of cryptology [1]. Then in 1990, Lamport and Leslie proposed Paxos algorithms with a high tolerance for errors while Haber and Stornetta concluded ledger-style data structures [2]. The two things indicated that the blockchain is just one step away. Finney improved Pow mechanism to full coverage of subsequent transport mechanisms [3]. At this time, all the theories and preparation for blockchain have done. In November of 2008, Satoshi Nakamoto published a paper, proposing a purely peer-to-peer version of electronic cash, which would allow online payments to be sent directly form one party to another party without any financial institution [4]. Subsequently in 2009, the Bitcoin was born, ushering in the era of Bitcoin. In 2013, the creator of Ethereum published White-Paper to launch the project of Ethereum, symbolling as the era 2.0 of Blockchain [5]. There are increasing types of cryptocurrencies, trading in the Internet actively and intending to lavish.

So, it is motivated for scholars to predict the price of the cryptocurrency, to make wise investment. Baek and Elbeck used the detrending ratio to test Bitcoin's relative volatility, and found that the Bitcoin market was highly speculative [6]. Cao found that all risks lead to a positive premium in the price of Bitcoin options with using a balanced valuation model of Bitcoin options [7]. Madan and Zhao in Stanford predicted the price of Bitcoin via binomial logistic regression, support vector machines, and random forests, comparing the accuracy of predicting the price of Bitcoin [8]. According to the available research, it can be concluded that machine learning which only the Bitcoin price index and conversion prices are accepted models that are trained have poor predictive performance. Saad used linear regression, random forest, and the long-short-term memory neural network model to make price predictions for Ethereum, and the results showed that long-short-term memory network model has high accuracy. Dian used the α-Sutte indicator method and the ARIMA model for forecasting [9]. Then, the conclusion is compared to NNAR, the α-Sutte indicator method has greater accuracy in predicting Bitcoin's transaction data. Franco Valencia (2019) used Twitter's relevant data and market transaction data as input features, the nerves were compared networks (NN), support vector machines (SVMs) and random forests (RF) for predictive performance. They

concluded that machine learning and sentiment analysis methods can be used to predict the price of cryptocurrencies, where neural networks perform better than other models, and it is feasible to use Twitter's relevant data to predict the price of certain cryptocurrencies [10]. Jaewook conducted a study of Bitcoin's historical price, comparing Bayesian neural networks with other prediction results of the lines models. They believe that Bayesian neural networks work best at predictions and can be explained high volatility of trading prices [11].

This study will use lightgbm, XGBoost and liner model to predict the future price of Bitcoin, Ethereum and Dogecoin. It can help the investors to have a deep understanding of the trend of the price and make the right choice in their investment by decreasing the risk of uncertainty. In this study, the Sec. 2 is the introduction of the sources and characteristics of the data, the models used to predict the price and the metrics to judge the accuracy of the models. Then there will be the results and discussion. On top of that, the limitation of this passage will be pointed out and some future steps will be provided.

2. Data & Method

The data is collected from cn.investing.com, downloading the daily data of Bitcoin, Ethereum and Dogecoin from 2020-01-01 to 2022-12-07, including close price, open price, high price, low price and the volume of trading coins. Given that Bitcoin and Ethereum have the largest trading volume and are the most typical cryptocurrencies which are traded in the market very actively, the passage chose the two to predict their price, in order to give some advice to the investors. With the aim of making the passage specific and comprehensive, Dogecoin was chosen to make comparison. The passage input the open price, high price, low price and the volume into models, outputting the close price. The test data is set as the last 30 official trading days and the rest data is set as the training data. The input values are close price, open price, high price, low price, and the volume of trading coins, while the output value is the close price.

To predict the data, this paper will use the lightgbm, XGBoost, and simple linear model to predict the future price of the three cryptocurrencies. To be specific, multiple linear regression is the simplest method in machine learning. For observational data, it calculates the best-fitted line by minimizing the sum of squared vertical deviations from each data point to the line. XGBoost is an ensemble learning algorithm that belongs to the category of boosting algorithms in 3 classes of commonly used ensemble methods (bagging, boosting, stacking). It is an additive model, the base model generally chooses the tree model, but other types of models such as logistic regression can also be selected. XGBoost belongs to the category of gradient boosting tree (GBDT) models, the basic idea of GBDT is to let the new base model (GBDT based on CART categorical regression tree) fit the bias of the previous model, thereby continuously reducing the bias of the additive model. Compared with the classic GBDT, XGBoost has made some improvements, leading in an important improvement in effect and performance. LightGBM is an improved GBDT (Gradient Boosting Decision Tree) model based on XGBoost. Compared to XGBoost, LightGBM has made the following optimizations.

To evaluate the model, the R-squared, MAPE and MSE were used to judge the accuracy of the regression. R-Squared, also known as coefficient of determination, also known as goodness-of-fit, reflects the degree to which the independent variable x explains the change in the dependent variable y. The closer it is to 1, the better the model fits, where the formula is

$$R^2(y, \hat{y}) = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \tag{1}$$

Here,

$$TSS = \sum_{i=0}^{n}(y_i - \bar{y})^2 \tag{2}$$

$$RSS = \sum_{i=0}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

$$ESS = \sum_{i=0}^{n}(\hat{y}_i - \bar{y})^2 \tag{4}$$

Mean absolute Error (MAE) is the expected value of absolute error loss, which can be calculated as:

$$MAE = \sum_{i=0}^{n} |y - \hat{y}_i| / n \tag{5}$$

Mean Square Error (MSE) corresponds to the expectation of squared error, which can be described as

$$MAE = \sum_{i=0}^{n}(y - \hat{y}_i)^2 / n \quad (6)$$

For MSE and MAE, the smaller the two are, the better the model predicts the data.

3. Results & Discussion

The overall prediction of the model is good. R-squared can reach more than 0.94, indicating a high degree of fit. Judging from the picture, each model can better predict the price trend, and the ups and downs are consistent. However, the accuracy of the predicted values needs to be improved. For XGBoost, in Bitcoin, the value of MSE is 153115.82, MAE is 288.86, and R-squared is 0.94. In Dogecoin, the value of MSE is close to 0, MAE is 0.00242, and R-squared is 0.97686. In Ethereum, MSE is 1652.55, MAE is 32.5, and R-squared is 0.94. For LightGBM, in Bitcoin, the value of MSE is 108952.09, MAE is 251.14, and R-squared is 0.94. In Gogecoin, the value of MSE is close to 0, MAE is 0.00235, and R-squared is 0.95838. In Ethereum, MSE is 1006.17, MAE is 25.75, and R-squared is 0.94. For Linear Model, in Bitcoin, the value of MSE is 42805.75, MAE is 141.79, and R-squared is 0.98. In Dogecoin, the value of MSE is close to 0, MAE is 0.00169, and R-squared is 0.98238. In Ethereum, MSE is 301.67, MAE is 12.46, and R-squared is 0.99. The results for Bitcoin, Ethereum and Dogecoin are summarized in Table. 1, Table. 2 and Table. 3. The price prediction results of the XGBoost, LightGBM and linear model for Bitcoin are shown in Fig. 1, Fig. 2, and Fig. 3, respectively. The price prediction results of the XGBoost, LightGBM and linear model for Ethereum are shown in Fig. 4, Fig. 5, and Fig. 6, respectively. The price prediction results of the XGBoost, LightGBM and linear model for Dogecoin are shown in Fig. 7, Fig. 8, and Fig. 9, respectively.

Table 1. Summary of the results of Bitcoin.

Bitcoin	MSE	MAE	R-squared
XGBoost	153115.82	288.86	0.94
LightGBM	108952.09	251.14	0.96
Linear Model	42805.75	141.79	0.98

Table 2. Summary of the results of Ethereum.

Ethereum	MSE	MAE	R-squared
XGBoost	1652.55	32.5	0.94
LightGBM	1006.17	25.75	0.96
Linear Model	301.67	12.46	0.99

Table 3. Summary of the results of Dogecoin.

Dogecoin	MSE	MAE	R-squared
XGBoost	1.04E-05	0.00242	0.97686
LightGBM	1.87E-05	0.00235	0.95838
Linear Model	7.90E-06	0.00169	0.98238

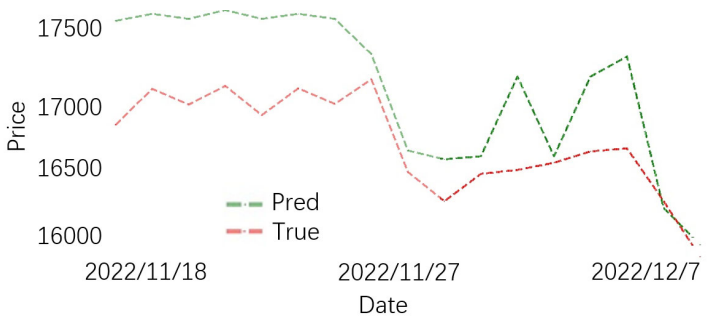

Fig. 1 Prediction of XGBoost for Bitcoin.

For Bitcoin, the Linear Model predicts best, with MSE, MAE and R-squared being the best performers of the three. The next best effect is LightGBM, and XGBoost performs the second. For Dogecoin, Linear Model predicted the best, followed by XGBoost and LightGBM the worst. For Ethereum, the Linear Model's R-squared reached a height of 0.99, a very high fit. This was followed by LightGBM performing the best, while XGBoost was the worst. For the three prediction models, one can see that the Linear Model is the best performer among the three cryptocurrencies' price predictions, while XGBoost is the worst. All R-squared values in the Linear Model can be as high as 0.95 or more, and the MSE and MAE values are much smaller than the other two models, indicating the model fits well and the prediction error is small.

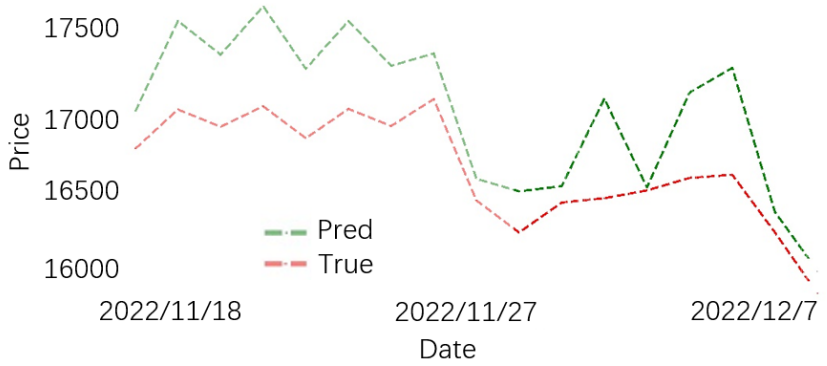

Fig. 2 Prediction of LightGBM for Bitcoin.

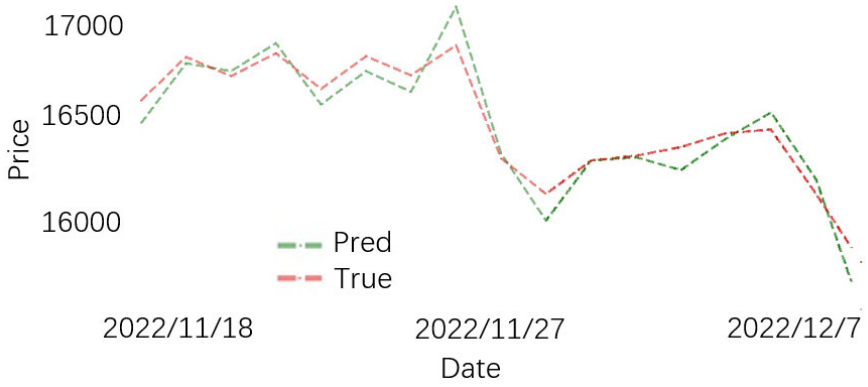

Fig. 3 Prediction of Linear Model for Bitcoin.

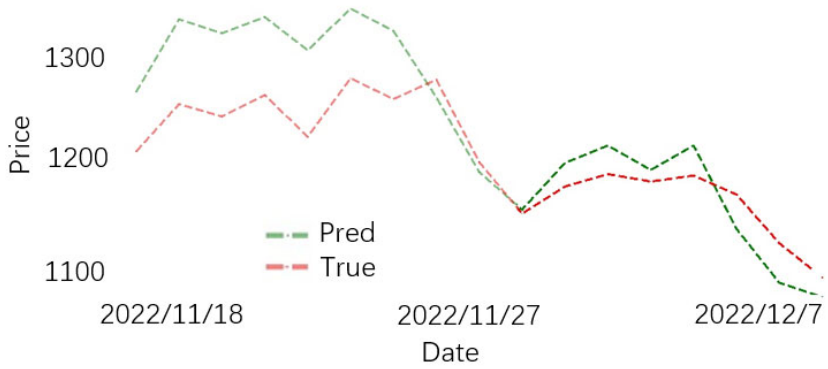

Fig. 4 Prediction of XGBoost for Ethereum.

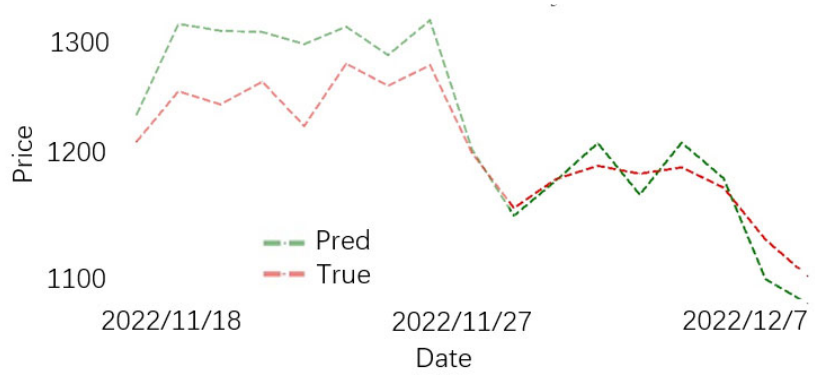

Fig. 5 Prediction of LightGBM for Ethereum.

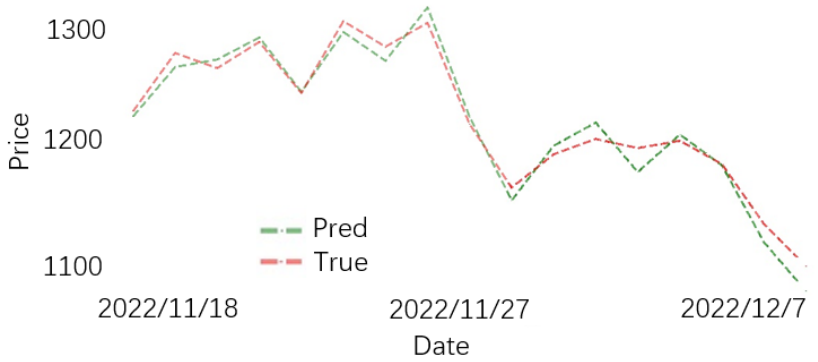

Fig. 6 Prediction of Linear Model for Ethereum.

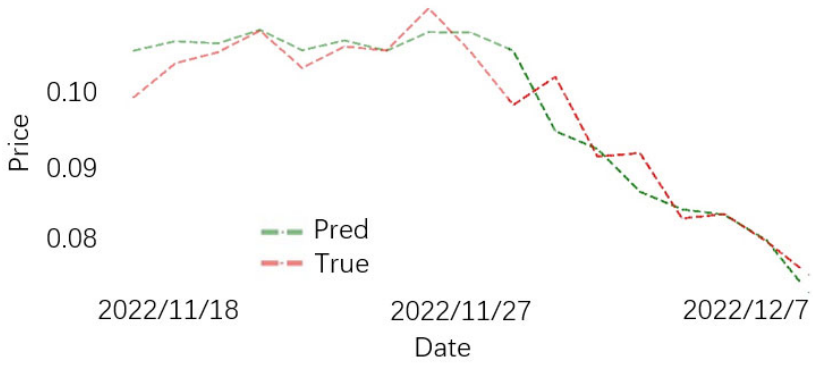

Fig. 7 Prediction of XGBoost for Dogecoin.

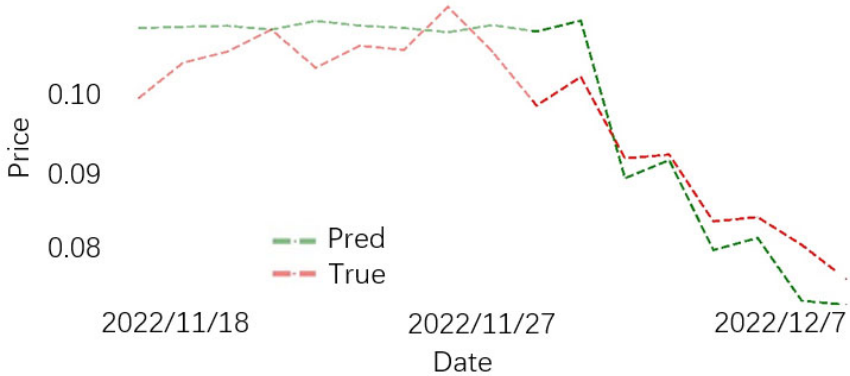

Fig. 8 Prediction of LightGBM for Dogecoin.

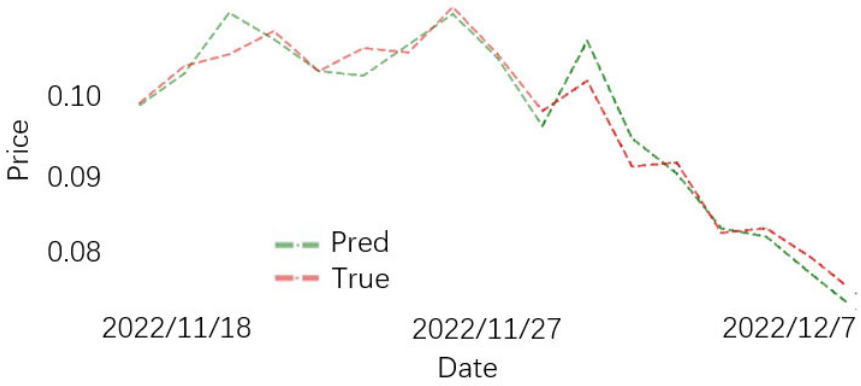

Fig. 9 Prediction of Linear Model for Dogecoin.

4. Limitations & Prospects

This paper uses only three machine learning models for cryptocurrency predictions, namely XGBoost, LightGBM, and Linear Model. In other studies, most scholars will use GARCH or other time-series-related models for time series forecasts such as cryptocurrency prices. At the same time, some scholars will use deep neural networks to predict prices. The three machine learning models in this article are simply regression prediction models, which are not enough in terms of innovation. In addition, the input variables for this article are only the opening price, closing price and market volume. In actual research and forecasting, other scholars will choose other factors (market volatility, consumer sentiment, macro market indexes, etc.). When more economically meaningful variables are added, forecasts are more accurate and easier to study and understand. In the future, there will be more statistical, mathematical and computer methods to make time series predictions, which will make the predictions more accurate and effective. Meanwhile, in the era of rapid computer development, cryptocurrencies have broader prospects, so their price predictions are more meaningful and valuable.

5. Conclusion

In summary, this paper investigates price prediction of cryptocurrencies based on machine learning. Specifically, this paper uses three models from machine learning, XGBoost, LightGBM and Linear Model. The opening price, closing price and market capacity of the three cryptocurrencies are brought into the model for training, and finally the predicted closing price is obtained according to the trained model. According to the analysis, all three models provide relatively accurate fitting and prediction of prices. The price action fits well and the forecast results are more accurate. In addition, according to the regression evaluation index and the picture, the Linear Model is the best performance of the three models. LightGBM performance is somewhere between Linear Model and XGBoost. Besides, Among the three cryptocurrencies, Dogecoin has the highest accuracy in its price prediction, while Bitcoin has the lowest price prediction. Nevertheless, the model used in this article is the most basic machine learning model with a small number of variables. In the future, scholars can try to make price predictions using a time series forecasting model such as GARCH. In the meantime, one can also consider adopting factor models to select factors with different economic significance to predict price trends, which makes the study more meaningful. Overall, these results offer a guideline for predicting price of cryptocurrencies by machine learning.

References

[1] Diffie W, Hellman M. New directions in cryptography. IEEE Transactions on Information Theory, 1976, 22(6):644-654

[2] Lamport L. The part-time parliament. ACM Transactions on Computer Systems, 1998, 16(2):133-169.

[3] Cao M, Celik B. Valuation of bitcoin options. Journal of Futures Markets,2021, 41(7): 1007-1026.

[4] Hal F. Reusable Proofs of Work. 2004, Retrieved from https://nakamotoinstitute.org/finney/rpow/.

[5] Vitalik B. A next-generation smart contract and decentralized application platform. 2013 Retrieved from: https://github.com/Ethereum/wiki/wiki/White-Paper, 2013.

[6] Haber S, Stornetta W S. How to Time-Stamp a Digital Document. Journal of Cryptology, 1991, 3(2).

[7] Madan S S, Zhao A. Automated Bitcoin trading via machine learning algorithms. Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep., 2015.

[8] Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 2017.

[9] Valencia F, Gómez-Espinosa A, Valdés-Aguirre B. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. Entropy, 2019, 21(6): 589.

[10] Utami S D, Saleh A A, Nuning K, et al. Forecasting Historical Data of Bitcoin using ARIMA and α-Sutte Indicator. Journal of Physics: Conference Series, 2018, 1028:012194.

[11] Jang H, Lee J. An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. IEEE Access, 2018, 6: 5427-5437.